

Northumbria Research Link

Citation: Anastasi, Elisa, MacArthur, Iain, Scotti, Mariela, Alvarez, Sonsiray, Giguère, Steeve and Vázquez-Boland, José A. (2016) Pangenome and Phylogenomic Analysis of the Pathogenic Actinobacterium *Rhodococcus equi*. *Genome Biology and Evolution*, 8 (10). pp. 3140-3148. ISSN 1759-6653

Published by: Oxford University Press

URL: <https://doi.org/10.1093/gbe/evw222> <<https://doi.org/10.1093/gbe/evw222>>

This version was downloaded from Northumbria Research Link:
<http://nrl.northumbria.ac.uk/id/eprint/44187/>

Northumbria University has developed Northumbria Research Link (NRL) to enable users to access the University's research output. Copyright © and moral rights for items on NRL are retained by the individual author(s) and/or other copyright owners. Single copies of full items can be reproduced, displayed or performed, and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided the authors, title and full bibliographic details are given, as well as a hyperlink and/or URL to the original metadata page. The content must not be changed in any way. Full items must not be sold commercially in any format or medium without formal permission of the copyright holder. The full policy is available online: <http://nrl.northumbria.ac.uk/policies.html>

This document may differ from the final, published version of the research and has been made available online in accordance with publisher policies. To read and/or cite from the published version of the research, please visit the publisher's website (a subscription may be required.)



**Northumbria
University**
NEWCASTLE



UniversityLibrary

Pangenome and Phylogenomic Analysis of the Pathogenic Actinobacterium *Rhodococcus equi*

Elisa Anastasi^{1,†}, Iain MacArthur^{1,†}, Mariela Scotti^{1,2}, Sonsiray Alvarez¹, Steeve Giguère³, and José A. Vázquez-Boland^{1,2,4,*}

¹Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

²Edinburgh Medical School (Biomedical Sciences), University of Edinburgh, Edinburgh, United Kingdom

³Department of Large Animal Medicine, University of Georgia, Georgia, USA

⁴Grupo de Patogenómica Bacteriana, Universidad de León, León, Spain

[†]These authors contributed equally to the study.

*Corresponding author: E-mail: v.boland@ed.ac.uk.

Accepted: September 2, 2016

Data deposition: This project has been deposited at GenBank under accessions LWTX000000000, LWIC000000000, LWB000000000, LWHS000000000, LWHT000000000, LWHU000000000, LWHV000000000, LWHW000000000, LWHX000000000, LWHY000000000, LWHZ000000000, LWIA000000000, LWIB000000000, LWTC000000000, LXFI000000000, LXFH000000000, LXFG000000000, LWHR000000000, LWTP000000000, LWTY000000000, LWTQ000000000, LWTR000000000, LWTS000000000, LWTT000000000, LWTU000000000, LWTV000000000, and LWTW000000000.

Abstract

We report a comparative study of 29 representative genomes of the animal pathogen *Rhodococcus equi*. The analyses showed that *R. equi* is genetically homogeneous and clonal, with a large core genome accounting for $\approx 80\%$ of an isolates' gene content. An open pangenome, even distribution of accessory genes among the isolates, and absence of significant core–genome recombination, indicated that gene gain/loss is a main driver of *R. equi* genome evolution. Traits previously predicted to be important in *R. equi* physiology, virulence and niche adaptation were part of the core genome. This included the lack of a phosphoenolpyruvate:carbohydrate transport system (PTS), unique among the rhodococci except for the closely related *Rhodococcus defluvi*, reflecting selective PTS gene loss in the *R. equi*–*R. defluvi* sublineage. Thought to be asaccharolytic, *rb*sCB and *glcP* non-PTS sugar permease homologues were identified in the core genome and, albeit inefficiently, *R. equi* utilized their putative substrates, ribose and (irregularly) glucose. There was no correlation between *R. equi* whole-genome phylogeny and host or geographical source, with evidence of global spread of genomovars. The distribution of host-associated virulence plasmid types was consistent with the exchange of the plasmids (and corresponding host shifts) across the *R. equi* population, and human infection being zoonotically acquired. Phylogenomic analyses demonstrated that *R. equi* occupies a central position in the *Rhodococcus* phylogeny, not supporting the recently proposed transfer of the species to a new genus.

Key words: *Rhodococcus equi*, pangenome analysis, comparative genomics, genome diversity and evolution, phylogenomics, *Corynebacteriales*, *Actinobacteria*.

Introduction

The soil-dwelling actinobacterium *Rhodococcus equi* is the causative agent of a purulent bronchopneumonic disease that affects foals in equine farms worldwide. In addition to horses, *R. equi* can also infect other animal species and is associated with severe opportunistic infections in immunocompromised people (Prescott 1991; von Bargen and Haas

2009; Vazquez-Boland et al. 2013). We previously reported the complete genome sequence of an equine isolate of *R. equi* (strain 103S). This work provided key information about the genome structure of the pathogen and the mechanisms of rhodococcal niche-adaptive genome plasticity and virulence evolution (Letek et al. 2010). Here we present the first comprehensive comparative genomic analysis of *R. equi*, involving

multiple isolates from different sources. Our new study provides insight into the core features, diversity, population structure and genome evolution of *R. equi*. It also clarifies the phylogenetic position of the species, repeatedly questioned based on equivocal 16S rDNA and numerical phenetic studies (Goodfellow et al. 1998; Gurtler et al. 2004; Jones and Goodfellow 2012; Jones et al. 2013b), unambiguously confirming *R. equi* is a *bona fide* member of the genus *Rhodococcus*.

Materials and Methods

Bacteria

The isolates sequenced in this study (supplementary table S1, Supplementary Material online) were selected to include at least two representatives from each of the seven major *R. equi* genogroups defined by Asel PFGE genotyping (Vazquez-Boland et al. 2008 and our unpublished data) plus the type strain of the species, DSM 20307^T (=ATCC 6939^T=ATCC 25729^T=NBRC 101255^T). Isolates of different animal sources (equine, bovine, porcine, ovine, human), geographical origin (13 countries) and host-associated virulence plasmid type carriage (pVAPA, pVAPB, pVAPN) (Takai et al. 2000; Letek et al. 2008; Valero-Rello et al. 2015) were analyzed.

Genome Sequencing and Analysis

Rhodococcus equi DNA was isolated from exponential cultures in BHI (OD₆₀₀ ≈ 1.0) using the GenElute™ kit (Sigma–Aldrich). Shotgun 101-bp pair-end DNA sequencing was performed at Beijing Genomics Institute (BGI, China) using TruSeq DNA PCR-Free Sample library preparation kit on Illumina HiSeq 2000 instruments. Strains 2274 to 2288 (supplementary table S1, Supplementary Material online) were sequenced at the genomics facility of the University of Georgia (USA) as previously described (Anastasi et al. 2015). Adaptors and low quality reads were trimmed using Scythe (<https://github.com/vsbuffalo/scythe>) and Sickel (<https://github.com/najoshi/sickle>), respectively, and assembled using SPAdes (Bankevich et al. 2012). Annotation was performed using Prokka V1.11 (Seemann 2014) and the complete 103S genome (Letek et al. 2010) as a reference. Pangenome analyses were performed using Get_Homologues V2.0 (Contreras-Moreira and Vinuesa 2013) with OrthoMCL clustering algorithm and 70% sequence identity–75% coverage as minimum BLASTp homology cutoff. Functional annotation was performed using BLASTKOALA (Kanehisa, et al. 2016) and the prokaryotes KEGG GENES search database.

Genome Diversity and Phylogenomic Analyses

Average nucleotide identity (ANI) was calculated using JSpecies (Richter and Rosselló-Móra 2009) with MUMmer alignment (ANIm) as described in Goris et al. (2007) (settings -X 150, -q -1, -F F, -e 1e-15, -a 2). *Rhodococcus equi* whole-

genome Maximum Likelihood (ML) phylogenetic reconstruction was performed with RealPhy (Bertels et al. 2014) using RAXML (Stamatakis 2014) for tree construction with the general time-reversible (GTR) model of nucleotide evolution and gamma distributed rate variation. The *Corynebacteriales* ML tree was constructed from alignments of concatenated conserved protein products using PhyloPhlan (Segata et al. 2013). Trees were graphed using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

Results and Discussion

Rhodococcus equi Is Genetically Homogeneous

Twenty-seven de novo determined *R. equi* whole-genome shotgun assemblies, the available draft genome of ATCC 33707, and the complete 103S genome (Letek, et al. 2010) were analyzed (supplementary table S1, Supplementary Material online). The average CDS number was 4,933 (range 4,525–5,325), similar to the gene content of the manually annotated 5.04-Mbp 103S genome (4,598) (Letek, et al. 2010). Mean G + C content was 68.77%, also similar to that previously determined for 103S (68.82%). The mean ANI value was 99.13% (range 98.86–99.28%), well above the consensus 95–96% threshold for prokaryotic species demarcation (Goris et al. 2007; Richter and Rosselló-Móra 2009; Kim et al. 2014). This corresponded to 100% 16S rDNA sequence identity (1,519 nt) across all the isolates.

In comparison, the ANI values with members of the two other main *Rhodococcus* lines of descent as defined based on 16 rDNA phylogenies (McMinn, et al. 2000; Jones and Goodfellow 2012), that is, the “*erythropolis*” clade (*R. erythropolis*, *R. jostii*, *R. opacus* and *R. fascians* included in the analysis) and the “*rhodochrous*” clade (*Rhodococcus rhodochrous*, *Rhodococcus rhodnii*, *Rhodococcus ruber*, and *Rhodococcus pyridinivorans* included in the analysis), were 72.27–74.58% and 68.55–75.15%, respectively. The ANI with the recently described *R. equi* close relative, *Rhodococcus defluvi* (strain Ca11^T) (Kampfer et al. 2014), was 82.96%. This corresponded to 16S rDNA identity values of 96–98% and 95–97% for representatives of the “*erythropolis*” and “*rhodochrous*” clades, respectively, and 99% for *R. defluvi*.

The above data correlate with a strong degree of genome similarity and synteny conservation in BLASTn alignments (fig. 1), indicating that *R. equi* is a genetically homogeneous species.

Rhodococcus equi Core and Pangenome

The core genome shared by all 29 *R. equi* strains comprises 3,858 homologous gene clusters (HGC) (fig. 2A), equivalent to 81.5% of the 103S genome or 78.2% of the average gene content of the analyzed isolates, reflecting a low degree of intraspecies genomic variability. A core genome size estimation plot starts plateauing at about 25–27 genomes (fig. 2B),

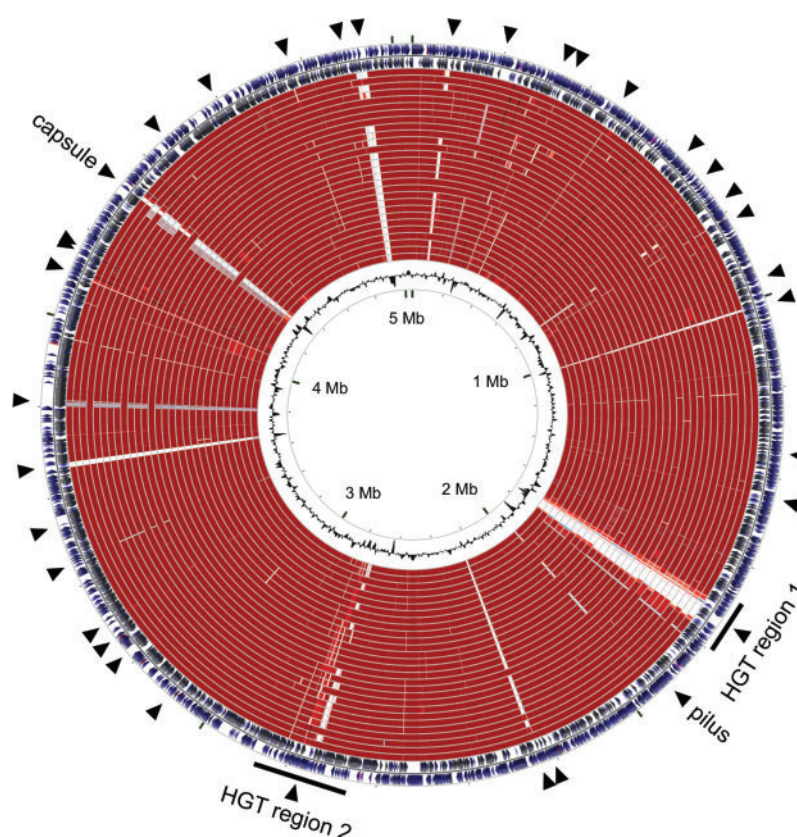


FIG. 1.—Genomic similarity of *Rhodococcus equi* isolates. BLASTn alignment of 28 draft genomes (inner rings) (supplementary table S1, Supplementary Material online) against the complete 103S chromosome (Letek, et al. 2010). Outermost two rings, 103S genes in forward and reverse strands. *E* value cutoff = 0.1. The predominant red colour in the aligned sequences indicates BLAST hit $\geq 98\%$ identity. Alignment gaps tend to coincide with regions of low G + C content in the 103S genome (innermost plot), many identified as HGT islands (arrowheads) by Alien_Hunter (Vernikos and Parkhill 2006). Drawn with CGViewer Comparison Tool (Grant et al. 2012).

indicating that the number of core genes is close to its maximum. The core genome contributes to 47.21% of the pangenome of the analyzed isolates ($n=8,174$ HGCs). About 35% of the pangenome is constituted by “cloud” HGCs, with a predominance of genes present in only one genome (fig. 2A), accounting for the species’ genome variability. This is consistent with the pangenome size plot, which increases almost linearly as new genomes are added (fig. 2B). The 4,316 HGC of the accessory pangenome are evenly distributed among the 29 isolates (fig. 2C), indicating a homogeneous pattern of genome evolution with similar rates of gene gain/loss processes across the *R. equi* population.

A KEGG functional classification showed similar overall distribution of categories between the core and the accessory genome, except for a proportional enrichment of genes involved in genetic information processing and nucleotide and cofactor/vitamins metabolism in the core genome, and in xenobiotic degradation, lipid metabolism and environmental information processing in the accessory genome (fig. 2D).

Specific Core Genome Features

We investigated whether specific traits identified in the 103S genome as potentially important for *R. equi* (Letek et al. 2010) belonged to the species’ core genome (supplementary table S2, Supplementary Material online). The absence of PTS sugar transport components (EI, HPr, EII complex/permeases) (Letek et al. 2010) was confirmed as a general feature of *R. equi*. This is likely due to gene loss because PTS components were present in all tested genomes from the other main lines of descent of the genus *Rhodococcus*. The PTS was also absent from the closely related *R. defluvii* Ca11^T (Kampfer et al. 2014), within the same terminal clade as *R. equi* in the *Rhodococcus* phylogeny (see below fig. 4), indicating that the gene loss event likely took place in the common ancestor of both species.

Two putative non-PTS sugar transporter genes were identified in the *R. equi* core genome: REQ19940-60 (103S annotation) encoding an RbsCB-like monosaccharide/ribose (xylose/arabinose) ATP-binding Cassette (ABC) transporter and cognate putative sugar kinase, and REQ20500 encoding

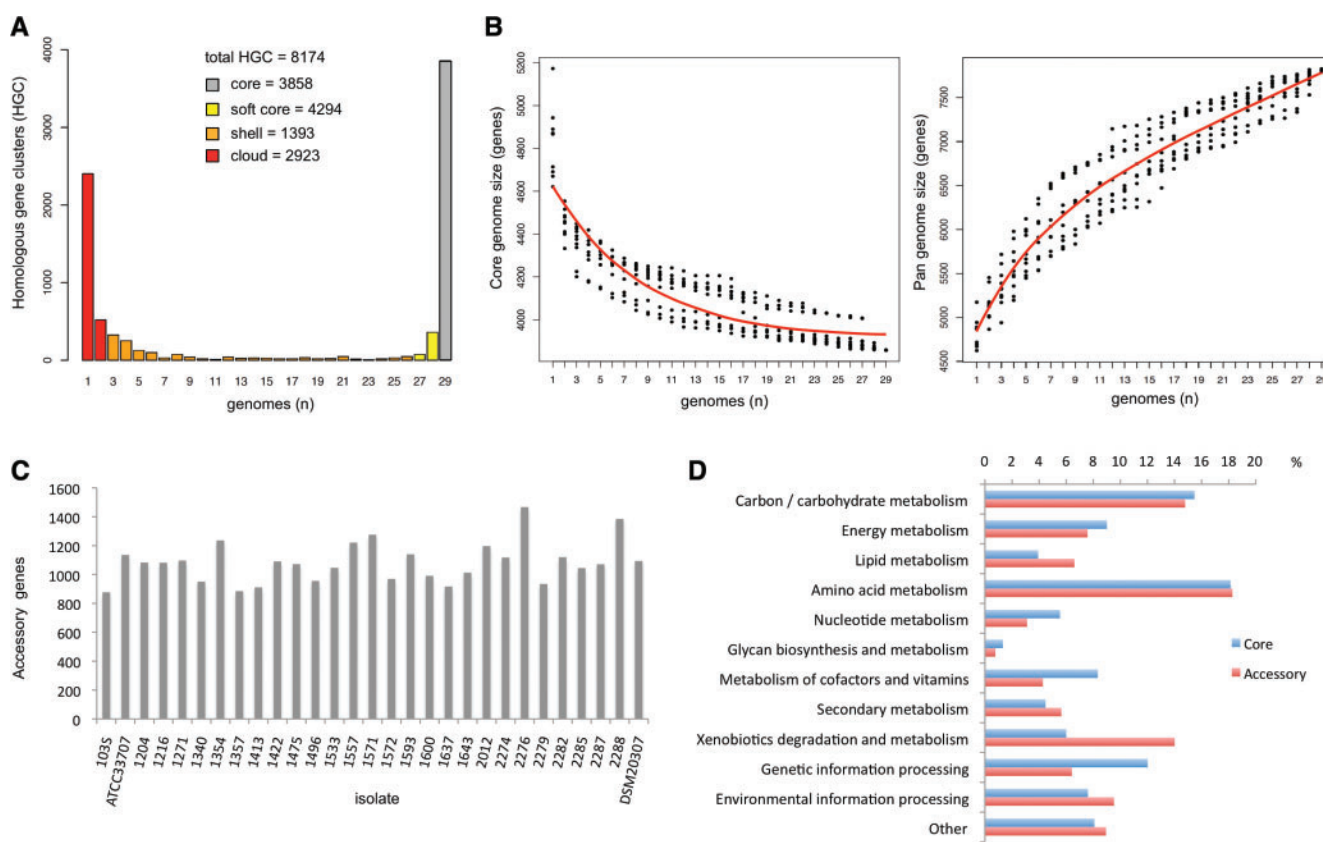


Fig. 2.—*Rhodococcus equi* core- and pangenome. (A) Pangenome distribution into strict core (present in 100% of isolates), soft-core (95% of isolates), cloud (≤ 2 genomes, cutoff defined as the class next to most populated noncore HGC) and shell (rest of HGCs). (B) Size estimation of core genome (left) and pangenome (right) by sequential sampling of n genomes in 10 random combinations using Tettelin exponential decay function fit (orthology threshold $\geq 50\%$ for C and S) (Tettelin et al. 2005). Analyses in (A) and (B) performed with Get_Homologues (Contreras-Moreira and Vinuesa 2013). (C) Distribution of accessory genes in *R. equi* isolates. The (manually curated) complete 103S genome (Letek et al. 2010) was subjected to automated annotation as a control; the lower number of accessory genes in the manually annotated 103S sequence ($n=667$) suggests that the gene content is overestimated in the draft genome sequences. (D) KEGG categories of core and accessory genome HGCs. Only 15.6% of the accessory genes could be categorized versus 45.2% for the core genome, indicating that the accessory genome is a source of functional innovation in *R. equi*.

a Major Facilitator Superfamily (MFS) permease similar to the *Streptomyces coelicolor* glucose transporter GlcP (van Wezel et al. 2005) (supplementary table S2, Supplementary Material online). Phenotype MicroArray (PMA) carbon source utilization tests (Bochner 2009) showed positive reactions for D-ribose, 2-deoxy-D-ribose, D-xylose (and its C'-2 carbon epimer L-lyxose), and D/L-arabinose (supplementary fig. S1A, Supplementary Material online). To exclude false positives due to abiotic dye reduction, growth curves were also performed in a chemically defined medium (mREMM, see supplementary fig. S1, Supplementary Material online, for details) using as a control L-lactate, a main carbon source for *R. equi* (Letek et al. 2010). Here only D-ribose consistently promoted *R. equi* growth, although after a protracted lag phase and to a lesser extent than L-lactate (supplementary fig. S1B, Supplementary Material online). In some experiments, delayed, weak growth was also observed with α -D-glucose (supplementary fig. S1B,

Supplementary Material online). Thus, while thought to be unable to metabolize carbohydrates (Letek et al. 2010), *R. equi* might utilize some sugars, albeit less efficiently than L-lactate and other preferred carbon sources (i.e., acetate and in general short- and long-chain monocarboxylates and fatty acids [Letek et al. 2010 and our unpublished observations]).

Virtually, all 103S loci potentially involved in tolerance to desiccation and oxidative stress, and thus important for *R. equi* survival in dry soil and transmission by aerosolized dust (Muscatello et al. 2007; Vazquez-Boland et al. 2013), were also found to be part of the core genome (supplementary table S2, Supplementary Material online). The same applies to the intrinsic resistome identified in 103S (9/10 β -lactamases, 5/5 aminoglycoside phosphotransferases and 4/4 multidrug efflux systems were conserved in all strains) (supplementary table S2, Supplementary Material online). Indeed, in vitro resistance to a number of antimicrobials, particularly

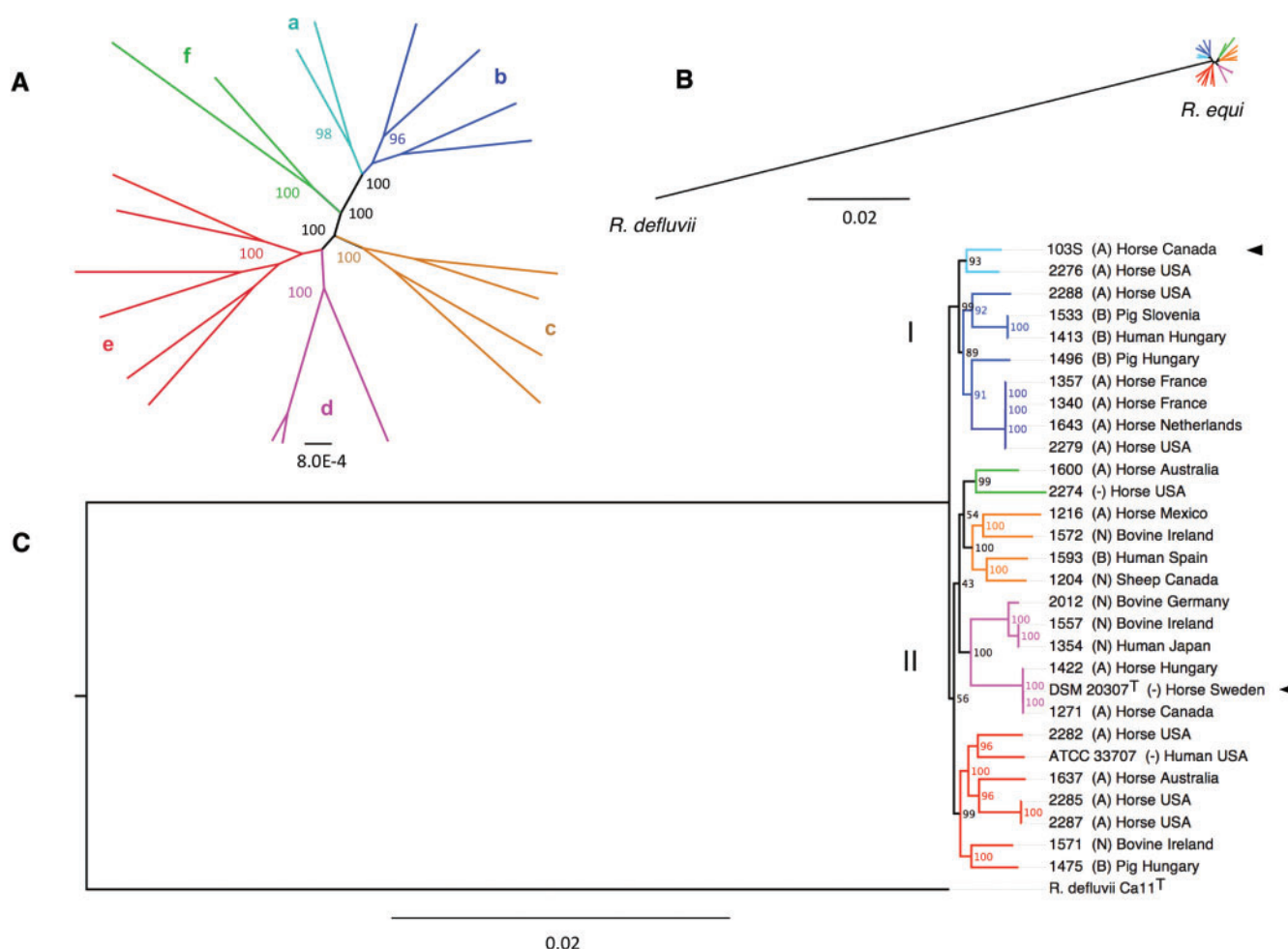


Fig. 3.—*Rhodococcus equi* core-genome phylogeny. ML trees inferred using RealPhy (Bertels et al. 2014). Nodes indicate bootstrap support from 500 replicates. Scale bars indicate substitutions per site. (A) Unrooted tree with *R. equi* subclades (a–f) highlighted in different colours. (B) Unrooted tree as in (A) including the genome of the closely related species *R. defluvii* Ca11^T (GenBank assembly accession GCA_000738775.1) to illustrate the tight clustering of *R. equi* strains (see also supplementary fig. S6, Supplementary Material online). (C) Same tree as in (B) rooted with *R. defluvii* Ca11^T. Tips show strain name, source of isolation (host, geographic origin) and plasmid type (confirmed by sequence analysis: A, equine pVAPA; B, porcine pVAPB; N, ruminant pVAPN; –, no plasmid; a detailed comparative analysis of the virulence plasmid genomes will be reported elsewhere). Arrowheads indicate the reference genome strain 103S (Letek et al. 2010) and the type strain of *R. equi* (DSM 20307^T). *Rhodococcus equi* isolates are split into two major lineages, I and II.

β -lactams and quinolones, has been observed in 103S (Letek et al. 2010) and reported in the literature for *R. equi* (Nordmann and Ronco 1992; Mascellino et al. 1994; Soriano et al. 1998; Makrai et al. 2000; Jacks et al. 2003; Jones and Goodfellow 2012).

All putative virulence-associated loci found in 103S, including those identified as HGT islands, that is, *mce2*, *srt1*, *srt2* and the pilus and capsule biosynthesis determinants (Letek et al. 2010), also belonged to the *R. equi* core genome (supplementary table S2 and fig. S2, Supplementary Material online). Two large HGT regions previously identified in 103S, likely generated by multiple horizontal gene acquisitions (Letek, et al. 2010), were also at least partially conserved in all isolates (fig. 1 and supplementary fig. S3, Supplementary Material

online). Since these genomic islands are all at the same chromosomal location in the genomes analyzed, the corresponding HGT events clearly occurred before *R. equi* diversification into sublineages (see below). The maintenance of a foreign DNA signature indicates a relatively recent acquisition, consistent with an evolutionarily young species.

Rhodococcus equi Core Genome Diversity and Population Structure

The species' phylogeny was reconstructed by analysis of single nucleotide polymorphisms in alignments of the draft genomes to the 103S reference genome. All *R. equi* isolates branched radially at a short distance (≈ 0.001 – 0.002 substitutions per

site between nodes of the major species' sublineages), denoting strong intraspecies genetic relatedness (fig. 3). The high degree of relatedness is most evident in a genomic ML tree including *R. defluvii* Ca11^T (fig. 3B and C), a species most closely related to *R. equi* according to 16S rDNA phylogenies (Kampfer et al. 2014) and whole genome comparisons (see above and [supplementary fig. S7, Supplementary Material online](#)). A recombination analysis showed no evidence of significant core-genome exchanges between strains ([supplementary fig. S4, Supplementary Material online](#)). Comparison of a parsimony tree based on a gene presence/absence matrix ([supplementary fig. S5, Supplementary Material online](#)) and the ML core-genome tree (fig. 3C) showed similar relationships between strains, indicating that the different *R. equi* sublineages tend to be associated with a similar accessory proteome composition. Overall, the above data is consistent with a clonal diversification pattern and a recent evolutionary origin for *R. equi*.

There was no obvious association between core-genome phylotypes and host source, whereas the latter was clearly linked with the host-associated plasmid type (fig. 3C). No correlation between genomic types and the geographical origin of the isolates was observed. This is illustrated by the equine strains DSM20307^T and PAM1271 or the bovine strains PAM1354 and PAM1557, which essentially share the same core and accessory genome while originating from Sweden and Canada, or Ireland and Japan, respectively (fig. 3C and [supplementary fig. S5, Supplementary Material online](#)).

Rhodococcus Phylogenomics

In a whole-genome phylogeny, the genus *Rhodococcus* appears as a distinct, well-defined monophyletic grouping of the *Corynebacteriales* (fig. 4 and [supplementary fig. S6, Supplementary Material online](#)). *Rhodococcus equi* isolates are clustered together in a *Rhodococcus* subclade (no. 3 or "equi" subclade) that contains two sister sublineages, one comprising *R. equi* and *R. defluvii* Ca11^T, confirming their close relatedness (Kampfer et al. 2014), and the other, *Rhodococcus triatomae* BKS15-14 and an unclassified isolate (fig. 4 and [supplementary fig. S6, Supplementary Material online](#)). Two other *Rhodococcus* subclades correspond to the 16S rDNA monophyletic groupings "rhodochrous" (subclade 1, with two sublineages: one encompassing *R. ruber*, another the type species of the genus, *R. rhodochrous*, and *Rhodococcus pyridinivorans*) and "erythropolis" (subclade 2, also with two sublineages: one with *R. opacus*, *R. jostii*, *Rhodococcus imtechensis* and *Rhodococcus wratislaviensis*, the other comprising *R. erythropolis* and *Rhodococcus qing-shengii*). Of note, subclades 2 ("erythropolis/jostii-opacus") and 3 ("equi") are sister lineages of a main *Rhodococcus* subdivision at the top of the genus tree (fig. 4 and [supplementary fig. S6, Supplementary Material online](#)). [Supplementary figure S7, Supplementary Material online](#), illustrates the genomic

relatedness between *R. equi* and representative members of *Rhodococcus* subclades 1, 2 and 3 in pairwise DNA sequence alignments.

Rhodococcus rhodnii LMG 5362 and *R. fascians* isolates define respectively two novel, more distantly related *Rhodococcus* subclades (nos. 4 and 5), the latter ("fascians") branching off at an early bifurcation in the genus phylogeny (fig. 4).

Rhodococcus and *Nocardia* form two clearly differentiated clades under a common node in the intermediate branchings of the *Corynebacteriales* (fig. 4 and [supplementary fig. S6, Supplementary Material online](#)). Both genera belong to a well-supported phyletic line that also comprises *Smaragdicoscus niigatensis* DSM44881^T, classified in the *Nocardiaceae* (as is *Rhodococcus*), as well as *Mycobacterium* spp. and *Amycolicoccus subflavus* (*Hoyosella subflava*) DQ53-9A1^T, classified in the *Mycobacteriaceae* (Ludwig et al. 2012). Another major *Corynebacteriales* phylogenomic subdivision is formed by members of the genera *Tsukamurella*, of the monogeneric *Tsukamurellaceae*, and *Gordonia* and *Williamsia*, in some taxonomies classified within the *Nocardiaceae* (Ludwig et al. 2012). The phylogenomic data therefore indicate that the *Nocardiaceae* taxon is polyphyletic and call for a reclassification of the genera *Rhodococcus*, *Nocardia* and *Smaragdicoscus* into a same (*Mycobacteriaceae*) family together with *Amycolicoccus* (*Hoyosella*) and *Mycobacterium*.

Conclusions

Our whole-genome comparative analyses show that *R. equi* is largely monomorphic, not supporting the commonly held view that *R. equi* is heterogeneous (McMinn et al. 2000; Jones and Goodfellow 2012; Jones et al. 2013b) and its isolates phylogenetically very diverse (Gurtler et al. 2004). The tendency of the core-genome sublineages to associate with a specific composition of the accessory genome and the lack of significant core-genome recombination indicate that *R. equi* evolution is primarily clonal. Although the accessory genome represents a relatively small fraction of an isolates' gene content (~20%), *R. equi* possesses an open pangenome that constitutes the basis of its genomic variability. The coincidence of the gaps in the genomic alignments with HGT islands in the complete 103S genome sequence indicates that lateral genetic exchanges have played a key role in the shaping of the *R. equi* accessory genome.

Our analyses show no evidence of phylogeographic correlation but instead of ample global circulation of genomotypes, probably linked to international livestock trade. The distribution of the host-associated virulence plasmid types in the *R. equi* phylogeny is consistent with the dynamic conjugal exchange of the plasmids across the *R. equi* population (Tripathi et al. 2012; Valero-Rello et al. 2015) and their key role in animal host tropism (Vazquez-Boland et al. 2013;

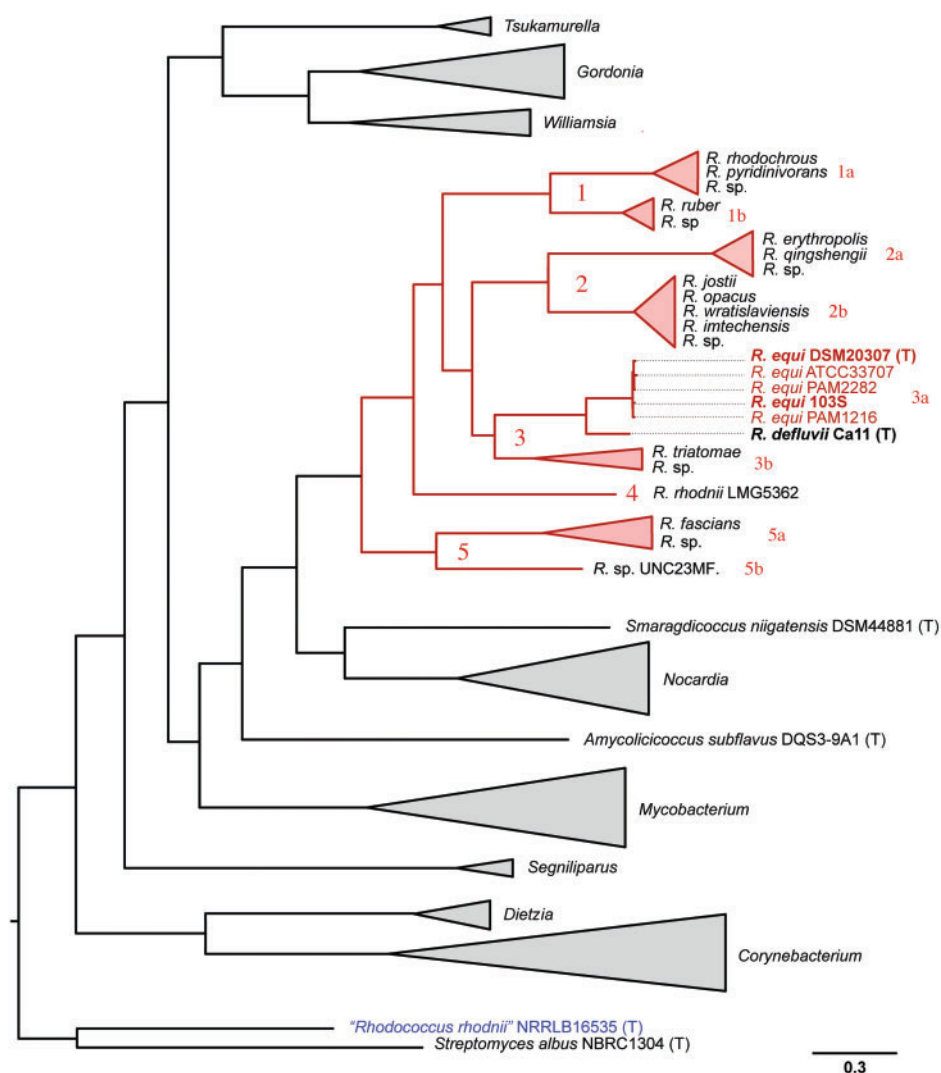


Fig. 4.—Whole-genome *Corynebacteriales* phylogeny. Constructed with PhyloPhlAn (Segata et al. 2013) using the genomes listed in [supplementary table S3, Supplementary Material](#) online. *Streptomyces albus* NBRC 1304^T was used as outgroup for tree rooting. Type strains are indicated by a T. All clades in the tree have been collapsed except the *Rhodococcus equi*–*R. defluvi* sublineage of *Rhodococcus* suclade 3. All nodes are strongly supported; see [supplementary figure S7, Supplementary Material](#) online, for a detailed tree with bootstrap values. *Rhodococcus* genus is in red, numbers designate major subclades (with letter suffix for sublineages). In blue, the genome of the type strain of *R. rhodnii* NRRL B-16535^T (GenBank assembly accession GCA_000720375.1) probably represents a case of strain mix-up or sequence mislabelling.

Valero-Rello et al. 2015). Strains sharing the same core and accessory genotome and virulence plasmid type were associated with both the corresponding adapted animal host and people (e.g., pVAPB-carrying 1413 and 1533 isolates, pVAPN-carrying 1354 and 1557 isolates) (fig. 3C and [supplementary fig. S5, Supplementary Material](#) online), strongly supporting that *R. equi* infection is zoonotically transmitted to humans (Ocampo-Sosa et al. 2007; Vazquez-Boland et al. 2013).

Further illustrating the remarkable uniformity of *R. equi*, virtually all major determinants predicted in 103S to be important for the species' biology, virulence and niche adaptation (Letek et al. 2010) were part of the core genome. This includes

the absence of a PTS and other specific metabolic traits such as the $\Delta thiC$ thiamin auxotrophic mutation or lactate utilization via a *lutABC* operon (Letek et al. 2010). These features may represent an adaptation to, and competitive advantage within the main saprophytic habitats of *R. equi*, manure-rich soil and the intestine (Muscatello, et al. 2007; Vazquez-Boland, et al. 2013), where microbially derived thiamine, and lactate and short-chain fatty acids produced by carbohydrate-fermenting microbiota, are presumably abundant.

Finally, our phylogenomic analyses resolve the lingering problem of *R. equi* taxonomy (Goodfellow et al. 1998; McMinn et al. 2000; Gurtler et al. 2004; Jones and

Goodfellow 2012; Ludwig et al 2012). It is evident from our data that *R. equi* is not at the periphery or outwith the genus *Rhodococcus*, closer to the *Nocardia*, as previously claimed (Goodfellow et al. 1998; McMinn et al. 2000; Jones et al. 2013b), but deeply embedded in the rhodococcal phylogeny. Indeed, the “*equi-defluvii-triatomae*” subclade (no. 3) forms with its sister “*erythropolis/jostii-opacus*” subclade (no. 2) a major monophyletic subdivision central to the genus *Rhodococcus* (fig. 4). In complete genome comparisons, *R. equi* 103S shows the same degree of pairwise homology to *R. erythropolis* PR4 and *R. jostii* RHA1 as these two subclade 2 members between themselves (Letek et al. 2010; Vazquez-Boland et al. 2013). This means that the recent proposal of transferring *R. equi* to a new genus “*Prescotella*”, with “*Prescotella equi*” as its sole species (Jones et al. 2013a, 2013b), would only be justified if new genera were also created for each *R. erythropolis* and *R. jostii*. Such an atomization of the genus *Rhodococcus* is unwarranted, because the rhodococci form, in the *Corynebacteriales* phylogenomic tree (see supplementary fig. S6, Supplementary Material online), a distinct monophyletic grouping equivalent in rank and diversity to other well-established genera, such as *Corynebacterium*, *Gordonia* or *Mycobacterium*.

Supplementary Material

Supplementary figures S1–S7 and tables S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We are greatly indebted to N. Fujita, National Institute of Technology and Evaluation (NITE), Japan, for making available the draft genome sequence of *R. rhodochrous* NBRC 16069^T. We also thank D. Lewis for her contribution in establishing our laboratory’s global *R. equi* isolate collection, and B. Contreras-Moreira for help with Get_Homologues software. This work was supported by the Horserace Betting Levy Board (grant nos. vet/prj/712 and vet/prj/753; to J.V.-B.) and core BBSRC funding from the Roslin Institute (BB/J004227/1). E.A. was supported by a BBSRC-funded Zoetis-sponsored CASE PhD studentship from the Centre for Infectious Diseases of the University of Edinburgh.

Literature Cited

- Anastasi E, et al. 2015. Novel transferable *erm*(46) determinant responsible for emerging macrolide resistance in *Rhodococcus equi*. *J Antimicrob Chemother.* 70:3184–3190.
- Bankevich A, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comp Biol.* 19:455–477.
- Bertels F, Silander OK, Pachkov M, Rainey PB, van Nimwegen E. 2014. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol Biol Evol.* 31:1077–1088.
- Bochner BR. 2009. Global phenotypic characterization of bacteria. *FEMS Microbiol Rev.* 33:191–205.
- Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol.* 79:7696–7701.
- Goodfellow M, Alderson G, Chun J. 1998. Rhodococcal systematics: problems and developments. *Antonie Van Leeuwenhoek* 74:3–20.
- Goris J, et al. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol.* 57:81–91.
- Grant JR, Arantes AS, Stothard P. 2012. Comparing thousands of circular genomes using the CGView Comparison Tool. *BMC Genomics* 13:202.
- Gurtler V, Mayall BC, Seviour R. 2004. Can whole genome analysis refine the taxonomy of the genus *Rhodococcus* FEMS Microbiol Rev. 28:377–403.
- Jacks SS, Giguère S, Nguyen A. 2003. In vitro susceptibilities of *Rhodococcus equi* and other common equine pathogens to azithromycin, clarithromycin, and 20 other antimicrobials. *Antimicrob Agents Chemother.* 47:1742–1745.
- Jones AL, Goodfellow M. 2012. Genus IV. *Rhodococcus*. In: Goodfellow, M., editors. *Bergey’s manual of systematic bacteriology*, Vol. 5, The Actinobacteria. New York: Springer. pp. 437–464.
- Jones AL, Sutcliffe IC, Goodfellow M. 2013a. Proposal to replace the illegitimate genus name *Prescottia* Jones et al. 2013 with the genus name *Prescottella* gen. nov. and to replace the illegitimate combination *Prescottia equi* Jones et al. 2013 with *Prescottella equi* comb. nov. *Antonie Van Leeuwenhoek* 103:1405–1407.
- Jones AL, Sutcliffe IC, Goodfellow M. 2013b. *Prescottia equi* gen. nov., comb. nov.: a new home for an old pathogen. *Antonie Van Leeuwenhoek* 103:655–671.
- Kampfer P, Dott W, Martin K, Glaeser SP. 2014. *Rhodococcus defluvii* sp. nov., isolated from wastewater of a bioreactor and formal proposal to reclassify [*Corynebacterium hoagii*] and *Rhodococcus equi* as *Rhodococcus hoagii* comb. nov. *Int J Syst Evol Microbiol.* 64:755–761.
- Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol.* 428:726–731.
- Kim M, Oh HS, Park SC, Chun J. 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int J Syst Evol Microbiol.* 64:346–351.
- Letek M, et al. 2008. Evolution of the *Rhodococcus equi* *vap* pathogenicity island seen through comparison of host-associated *vapA* and *vapB* virulence plasmids. *J Bacteriol.* 190:5797–5805.
- Letek M, et al. 2010. The genome of a pathogenic *Rhodococcus*: cooptive virulence underpinned by key gene acquisitions. *PLoS Genet.* 6:e1001145.
- Ludwig W, et al. 2012. Road map of the phylum Actinobacteria. In: Goodfellow M, et al., editors. *Bergey’s Manual of Systematic Bacteriology*, Vol. 5, The Actinobacteria. New York: Springer. pp. 1–28.
- Makrai L, et al. 2000. Characterisation of *Rhodococcus equi* strains isolated from foals and from immunocompromised human patients. *Acta Vet Hung.* 48:253–259.
- Mascellino MT, Iona E, Ponzo R, Mastroianni CM, Delia S. 1994. Infections due to *Rhodococcus equi* in three HIV-infected patients: microbiological findings and antibiotic susceptibility. *Int J Clin Pharmacol Res.* 14:157–163.
- McMinn EJ, Alderson G, Dodson HI, Goodfellow M, Ward AC. 2000. Genomic and phenomic differentiation of *Rhodococcus equi* and related strains. *Antonie Van Leeuwenhoek* 78:331–340.
- Muscatello G, et al. 2007. *Rhodococcus equi* infection in foals: the science of ‘rattles’. *Equine Vet J.* 39:470–478.
- Nordmann P, Ronco E. 1992. In-vitro antimicrobial susceptibility of *Rhodococcus equi*. *J Antimicrob Chemother.* 29:383–393.

- Ocampo-Sosa AA, et al. 2007. Molecular epidemiology of *Rhodococcus equi* based on *traA*, *vapA*, and *vapB* virulence plasmid markers. *J Infect Dis.* 196:763–769.
- Prescott JF. 1991. *Rhodococcus equi*: an animal and human pathogen. *Clin Microbiol Rev.* 4:20–34.
- Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A.* 106:19126–19131.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.
- Segata N, Bornigen D, Morgan XC, Huttenhower C. 2013. PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun.* 4:2304.
- Soriano F, Fernandez-Roblas R, Calvo R, Garcia-Calvo G. 1998. *In vitro* susceptibilities of aerobic and facultative non-spore-forming gram-positive bacilli to HMR 3647 (RU 66647) and 14 other antimicrobials. *Antimicrob Agents Chemother.* 42:1028–1033.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Takai H, Hines SA, Sekizaki T, Nicholson VM, Alperin DA, Osaki M, Takamatsu D, Nakamura M, Suzuki K, Ogino N, Kakuda T, Dan H, Prescott JF. 2000. DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103. *Infect. Immun.* 68:6840–6847.
- Tettelin H, et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci U S A.* 102:13950–13955.
- Tripathi VN, Harding WC, Willingham-Lane JM, Hondalus MK. 2012. Conjugal transfer of a virulence plasmid in the opportunistic intracellular actinomycete *Rhodococcus equi*. *J Bacteriol.* 194:6790–6801.
- Valero-Rello A, et al. 2015. An invertron-like linear plasmid mediates intracellular survival and virulence in bovine isolates of *Rhodococcus equi*. *Infect Immun.* 83:2725–2737.
- van Wezel GP, et al. 2005. GlcP constitutes the major glucose uptake system of *Streptomyces coelicolor* A3(2). *Mol Microbiol.* 55:624–636.
- Vazquez-Boland JA, et al. 2008. Epidemiology and evolution of *R. equi*: new insights from molecular studies. In *Proceedings of the 4th Havemeyer Workshop on Rhodococcus equi*, p. 74. Edinburgh, UK.
- Vazquez-Boland JA, et al. 2013. *Rhodococcus equi*: the many facets of a pathogenic actinomycete. *Vet Microbiol.* 167:9–33.
- Vernikos GS, Parkhill J. 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics* 22:2196–2203.
- von Borgen K, Haas A. 2009. Molecular and infection biology of the horse pathogen *Rhodococcus equi*. *FEMS Microbiol Rev.* 33:870–891.

Associate editor: Howard Ochman